## **Introduction**

The inheritance of two genomic segments that are identical by descent resultant from an inbreeding event of related parents can generate runs of homozygosity (ROH) in their offspring. With increased availability of high density genomic data such a SNP-chips and full sequence data, in recent years the use of ROH in research has increased exponentially ((Pryce et al., 2014)ref). The main use of ROH to date being as an estimator of inbreeding. The ROH derived inbreeding coefficient FROH maintains advantages over traditional methods of inbreeding coefficients such as Fgrm and Fped. The main advantages being the absence of the need for a pedigree (as in Fped) and the ability to distinguish between homozygous alleles that are identical by decent i.e. a product of inbreeding, rather than identical by state i.e. the same allele mutation occurring through different mutations (Pryce et al., 2014, Keller et al., 2011). A number of studies have used FROH to infer inbreeding depression (reduced fitness occurring as a result of inbreeding) in a variety of organisms, from humans to dogs and cattle (Bjelland et al., 2013, Keller et al., 2012, Pryce et al., 2014). Following this, breeds or populations can be identified which may be at risk of recessive diseases or an overall reduction in fitness (Bjelland et al., 2013, Pryce et al., 2014).

Previous studies have shown that the length, abundance and genomic location of ROHs varies considerably in a population and can be affected by a number of related processes, including: selection, recombination, and population history (Kardos et al., 2017). Firstly, an allele under strong positive selection is expected to increase in incidence in the population which may result in increased homozygosity at the site and eventually lead to an area where ROH are common in the population. This concept has been widely used to identify signatures of selection in artificially selected organisms. By searching for areas in the genome where ROH are unusually common, often called ROH hotspots or ROH islands, it is possible to identify genomic regions of interest and even identify genes under selection. For example, genes associated with lactation and milk yield have been found to be present in dairy cattle, in regions where more than 50% of the population harboured a ROH (Peripolli et al., 2018). Once identified these regions can thence be used to aid genetic improvement of livestock.

ROH size and position has also been shown to correlate with recombination rate. Previous studies show that ROH hotspots also tend to be sites of lower recombination rate (Pemberton et al., 2012). In low recombination regions with high linkage disequilibrium long haplotypes are rarely broken up through meiosis, meaning they are more likely to come together in a ROH, forming a ROH hotspot. Equally, in high recombination regions, long haplotypes are broken down, resulting in fewer and shorter ROH. However, selection and recombination rate should not be thought of as mutually exclusive. Selection may interact with recombination rate to influence ROH distribution (Kardos et al., 2017). Selection can reduce genetic variation and Ne at closely linked loci (called genetic hitchhiking or a selective sweep), and the extent to which is dependent on the selection coefficient and local recombination rate (Charlesworth, 2009). As such, in an area of low recombination rate, a site under positive selection is more likely to have an extended region of linked loci, which may then result in a ROH. Therefore, both selection and recombination rate may be influencing the formation of a ROH hotspot.

An additional factor that can influence ROH distribution is population history. Intuitively, population history can directly influence ROH abundance, as the more inbred a population is the more ROH will be present in the population. For example, populations of both wild and domesticated pigs having undergone a known population bottleneck have higher incidence of ROHs than populations not having a bottleneck (Bosse et al., 2012). It was further shown by Bosse et al. (2012) that animals from the same populations tended to have similar ROH distribution patterns reflecting a shared demography. This implies that population history can have dramatic effect on ROH location, and shared population history events can result in a common ROH pattern in all individuals which experience the same history. Factors such as drift and selection will also play an important role in shaping the ROH landscape following such an event. For example, following a population bottleneck, genetic drift may result in certain alleles becoming fixed by chance, increasing the likelihood of generating ROHs at said site following generations of inbreeding. In addition, sites of past selection can result in common ROH hotspots between populations (Pemberton et al., 2012). As in Grilz-Seger et al. (2019) a gene associated with coat colour suspected to be under positive artificial selection in horses was shown to be in a ROH hotspot common between breeds which shared the same founders and ancestors. On the other hand, population specific ROH hotspots can indicate unique sites of selection (Grilz-Seger et al., 2019, Pemberton et al., 2012). It is also important to note the interactions between recombination rate and population history. Species or populations which are closely related will have similar recombination rate patterns across the genome, which will again result in more similar distributions of ROHs (Nothnagel et al., 2010).

Together the factors discussed above: recombination rate, selection and population history all play interacting roles in shaping the genomic distributions of ROH in a population. Here we aim to investigate the influence of each of these factors on ROH distribution in a wild population of red deer inhabiting the island of Rum, Scotland. We use >35,000 genome-wide autosomal SNPs to search for ROH in the population over a 40-year-period. We also use an additional population from mainland Scotland to compare inbreeding level and ROH distribution in different populations. In addition, a linkage map for red deer enabled the quantification of genomic recombination allowing for the assessment of recombination rate effects on ROH. We also ran forward simulations based on known history of the population of red deer on Rum to investigate the effects of selection, recombination and population history on the distribution of ROHs.

## **Methods**

***Study populations***

The red deer population inhabiting the north block of the Isle of Rum, Scotland (57°0’N, 6°20’W) has been studied at an individual level since 1971 (Clutton-Brock et al.1982), and is the main study population for this project. DNA was extracted from ear punches from calves captured soon after birth or darted adults, post-mortem tissue or cast antlers (See Huisman et al. 2016 for full details). DNA samples were genotyped at 50,541 attempted SNP loci on the Cervine Illumina BeadChip (Brauning et al. 2015 and see Johnston et al. 2017 for full details). SNP quality control was as follows: MAF > 0.01, ID genotyping success > 0.99, SNP genotyping success > 0.99, after which 35,132 autosomal SNPs and 3046 individuals were retained for analysis.

A mainland population of red deer from Kintyre, Scotland was compared with the Rum deer. Details of sample collection and DNA extraction can be found in McFarlane et al. (2019). Samples were genotyped and quality controlled as above, leading to 34,673 autosomal SNPs and 157 individuals being retained for analysis. These individuals were originally genotyped as part of a study of red x sika (*Cervus nippon*) hybridisation and were pure red deer according to that study (McFarlane et al. 2019).

***Calling runs of homozygosity***

We searched for runs of homozygosity in each genotyped individual using *homozyg* within the Plink v1.90 software (Purcell et al. 2007). We used two different estimates of marker position to call ROH: recombination distance in centimorgans (cM) based on direct estimates of recombination in the Rum red deer pedigree (Johnston ref) and physical distances in basepairs (BP) . The following parameters were used to identify ROH in both cases: minimum number of SNPs in a ROH of 40, minimum length of a ROH of 2500 Kb/cM, 1 SNP expected every 70 Kb /M, sliding window size of 35 SNPs, 4 missing calls allowed in a window, 0 heterozygotes allowed in a window, and minor allele frequency 0.01.

***ROH size classes and inbreeding coefficients***

Identified ROH were classified into three length categories to reflect distant inbreeding events, more recent and very recent inbreeding events, based on the relationship between ROH length and generations since a common ancestor, 𝑙 = 100 2𝑔 𝑐𝑀 (Thompson 2013). The categories were as follows, with the corresponding mean number of generations to the common ancestor in brackets:   
Short ROH: 2.5mbp (~20 g) – 5mbp (10 g),   
Medium ROH: 5Mbp (10g)– 16Mbp (~3 g),   
Long ROH > 16Mbp (≤3 g)

A genome-wide inbreeding coefficient, FROH for each individual was calculated using the sum of Kb or cM in ROH across all autosomes divided by the total Kb or cM of autosomes estimated at 2,495,700Kb (Johnston *et al.* 2017). We also calculated a ROH length-specific FROH as the proportion of the genome made up by each ROH length group (short, medium, long as defined above).

Here I think you should have a paragraph saying that we also estimated genome-wide inbreeding coefficients using the pedigree (cite Wright 1930) and a SNP-by-SNP estimator Fgrm (cite Yang). We then correlated these different estimates to confirm that FROH is indeed a measure on inbreeding. Scope for comparing the different FROHs too? I would find it interesting if the best correlations were with (say) FROHlong.***ROH Hotspots and coldspots***

We estimated the proportion of individuals (*P*, within either Rum or Kintyre) at which each SNP was in a ROH. In the course of this activity we found that *P* was positively correlated with the density of SNPs in a 1500kb/cM window (correlation for Rum population using cM as SNP position, Pearson’s correlation r = 0.58). This is probably an artefact of the density of SNPs in a region and the Plink algorithm for calling ROH. A density of 23 SNPs per 1500kb/cM window with a 100kb/cM sliding window was chosen as the minimum acceptable SNP density because this was where the relationship saturated and after this threshold the correlation is greatly reduced (Pearsons correlation: 0.2, Supplementary information). All SNPs within windows that fell below this density were discarded from the analysis of hotspots and coldspots. Additional QC for this analysis removed the first and last 40 SNPs from each linkage group to account for the fact that fewer ROH will be called in these regions as a ROH cannot span past the extremities of a linkage group. These processes resulted in 25,798 SNPs remaining using the linkage map and 26,318 SNPs using the physical map for the Rum population. In the Kintyre population 25,194 SNPs were retained using the linkage map and 25,638 SNPs using physical map. The top 1% *P* regions were classed ROH hotspots and the bottom 1% *P* were classed as coldspots (ref?). This method was used as it offers a simple population-specific evaluation of hotspots.

***Haplotype diversity***

Phased haplotypes for each individual in the Rum dataset were obtained using Alphapeel (ref) resulting in two haplotypes (one maternal and one paternal) per individual across all linkage group (Peters et al. in prep). Each linkage group was split into 10 SNP window haplotypes and any haplotypes with missing calls within them were removed (many of these instances were long uncalled haplotypes). The number of unique haplotypes in each 10 SNP window and the number of occurrences of each, was counted. A haplotype diversity measure was based on Simpson’s diversity index for measure of biodiversity:

Where *n* = the number of occurrences of each haplotype and *N* = total number of occurrences of all haplotypes and range is 0 (low diversity) to 1 (high diversity).

***Simulations***

Simulations were carried out in SLiM (Messer) to test the effect of selection, recombination rate and population history on the distribution of ROH across a simulated 100Mb chromosome. Each simulation was run 25 times. We used three different population history scenarios: one reflecting the Rum population, one reflecting the mainland population and one with a more severe population bottleneck than the Rum population. All scenarios started with an effective population size of 2000 set to run for 20,000 generations as a burn-in (10\*Ne). Based on the estimation of Ne = ~1/10 population size with ~200,000 individuals on the mainland. The Rum population scenario then drops to Ne = 100 at generation 20,000 to simulate the introduction of the deer to the isle of Rum ~160 years ago. The simulation ends at generation 20,023 (simulated present day) and outputs the current 100 individuals. The more severe bottleneck scenario drops Ne = 10 at generation 20,000, and then Ne = 100 at generation 20,005. As above the simulation ends at generation 20,023 and outputs 100 individuals. The mainland population scenario maintains Ne = 2000 until generations 20,023 where 2000 individuals were output.

Five different models were tested for each population history scenario: Neutral, Selection, Varied recombination rate, varied recombination rates and selection, and finally, varied recombination rates with higher selection coefficients. All models had a mutation rate of 1e-8. The neutral model was set to have a constant recombination rate across the chromosome at 1.038 cM/Mb based on estimates from the linkage map (Johnston) and every mutation was set as neutral at s=0.   
The selection model incorporated: neutral mutations (fixed s=0), slightly beneficial mutations (mean s=0.0001, exponential distribution), beneficial mutations (mean s=0.001, exponential distribution), slightly deleterious mutations (mean s=-0.001, gamma distribution shape parameter 0.2) and deleterious mutations (mean s=-0.01, exponential distribution) occurring in the ratio 4:1:1:1:1 respectively. Selection coefficients for each mutation were drawn from a distribution with a mean selection coefficient both stated above.   
The recombination model varied recombination rate across the chromosome based on estimates of recombination rates across all autosomes from Johnston et al. We split the 100Mb chromosome into 10 10Mb regions and averaged the relative position recombination rate across all autosomes. For regions 1-10 the recombination rates were as follows: 1.75 cM/Mb, 1.23 cM/Mb, 0.89 cM/Mb, 0.81 cM/Mb, 0.74 cM/Mb, 0.80 cM/Mb, 0.87 cM/Mb, 1.05 cM/Mb, 1.20 cM/Mb, 0.67 cM/Mb. The model including both recombination and selection did so as described in the previous two models. The model with recombination and higher selection coefficients altered the selection coefficients by a factor of five. As such, neutral mutations (fixed s=0), slightly beneficial mutations (mean s=0.0005, exponential distribution), beneficial mutations (mean s=0.005, exponential distribution), slightly deleterious mutations (mean s=-0.005, gamma distribution shape parameter 0.2) and deleterious mutations (mean s=-0.05, exponential distribution).

## **Results**

Results for both the linkage and physical map positions will be reported here, enabling comparison between the two marker positions and their effect on discovered ROH. Across all 3046 individuals in the Rum deer we found 55,414 ROH using the physical map and 45,814 ROH using genetic map. The majority of ROH we found were short and medium ROH, whereas few large ROH were found, consistent between maps and populations, Table 1, supplementary. Additionally, short and medium ROH covered on average between 1-2.5% of an individual’s genome, with medium ROH covering slightly more as a result of size classifications. Long ROH covered less on average but this was higher when using the physical map. Together this indicates that here the majority of individuals in a population have several short and medium ROH whereas very few individuals have many long ROH. Those few, often inbred, individuals with many long ROH can accumulate up to 20% of an individual’s genome, which were more prevalent in the Rum population than the Kintyre population, seen in the outliers of Sup Fig.

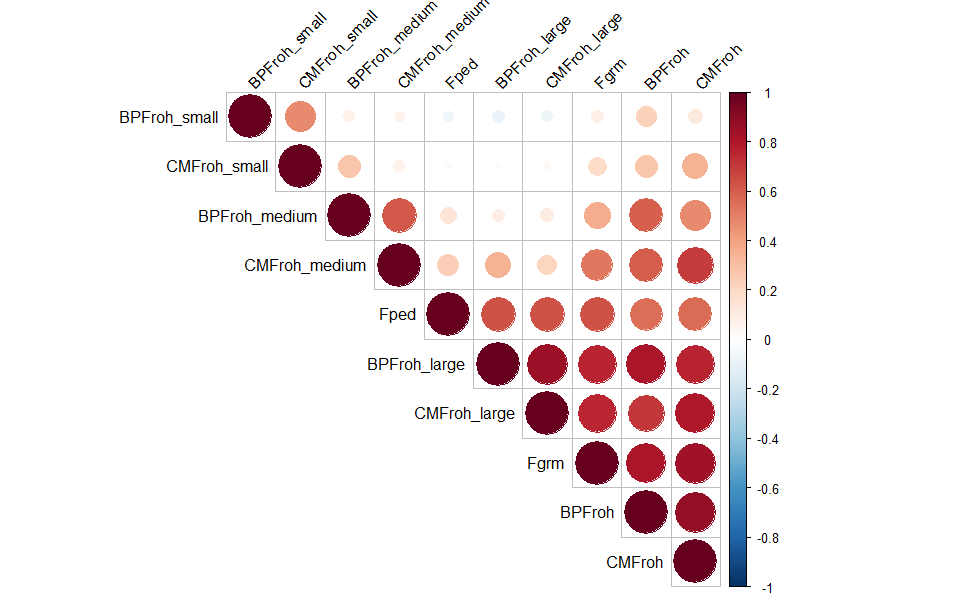
**Table 1** – Table showing the percentage of total ROH, mean number per individual and the average percentage of the genome in a ROH for each ROH category using genetic and physical map positions for the Rum population . ROH categories include: short (2.5mbp – 5mbp), medium (5Mbp – 16Mbp), long (> 16Mbp) and all ROH.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | ROH category | % of total ROH | Mean number per id | Ave % of genome in ROH |
| Genetic map positions | ROH short | 55.8 | 8.4 | 1.21 |
| ROH medium | 40.4 | 6.08 | 1.94 |
| ROH long | 3.8 | 0.56 | 0.54 |
| ROH all | N/A | 15.04 | 3.7 |
| Physical map positions | ROH short | 52.6 | 9.5 | 1.37 |
| ROH medium | 41.9 | 7.6 | 2.5 |
| ROH long | 5.5 | 1 | 1 |
| ROH all | N/A | 18.2 | 4.88 |

**Table 2** – Comparison of mean, minimum and maximum number of ROH per individual, mean inbreeding coefficient and maximum individual inbreeding coefficient between two populations and two map positions.

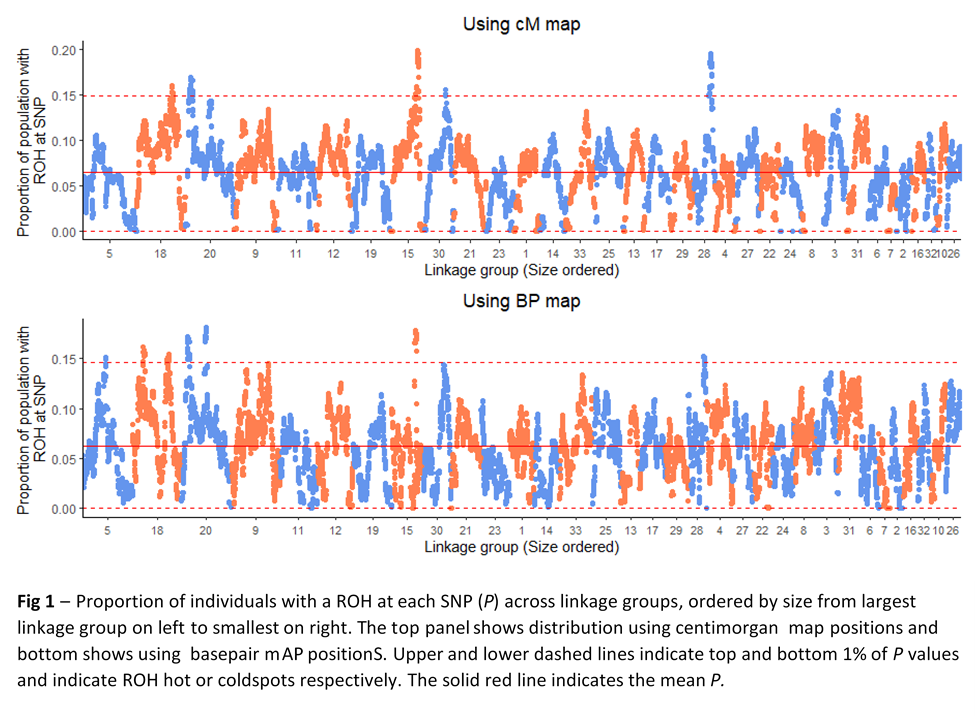
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Mean number ROH per id | Min ROH | Max ROH | Mean FROH | Max FROH |
| Rum population  (Genetic map) | 15.04 | 0 | 34 | 0.037 | 0.193 |
| Kintyre pop  (Genetic map) | 7.82 | 0 | 22 | 0.019 | 0.107 |
| Rum population  (Physical map) | 18.2 | 0 | 36 | 0.0488 | 0.261 |

The rum population had a higher inbreeding coefficient than the Kintyre populations and overall more ROH. The mean and maximum population inbreeding coefficients are higher using physical map positions for both populations, Table 2. Using the Rum population physical and genetic map inbreeding coefficients were highly correlated, as was the other genetic inbreeding coefficient Fgrm. The pedigree based method – Fped – was less correlated with other methods of full inbreeding coefficients. Small and medium FROH were poorly correlated with other methods, whereas large FROH were well correlated with other methods of full inbreeding coefficients.



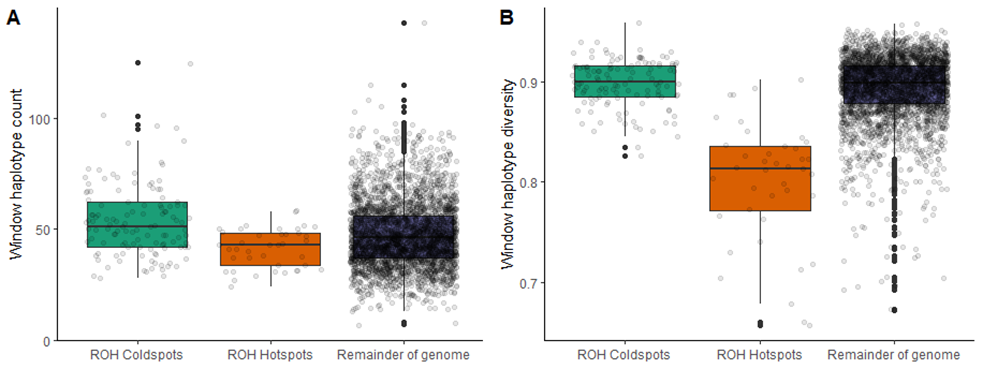
***ROH hotspots***

As demonstrated above we find that the marker positions used (cM or bp position) affected the number of ROH found in the population, we will now explore this effect on genomic distributions of ROH. In Rum deer using the genetic map positions the mean proportion of individuals with a ROH at a particular SNP (*P*) was 0.068 ± 0.0002 (SE) with *P* varying between and within linkage groups as seen in Fig.2. The threshold for the top 1% of *P* values occurred when >14.9% of individuals in the population had a ROH at a particular SNP. There were 257 SNPs that fell into this ROH hotspot category located in five regions across five chromosomes: 15, 18, 20, 28 and 30 (Fig.2). The bottom 1% *P* fell at zero individuals having a ROH at a SNP and 549 SNPs were categorised as a coldspot in this way. Coldspots are harder to visualise in Fig. 1A due to quality controls but occur on 13 different linkage groups and mainly at the end and beginning of linkage groups.   
In contrast, using the physical map positions the mean P was and the top 1% P occurred when >14% of individuals had a ROH at a SNP. There were seven hotspot regions on five linkage groups: five, 15, 18, 20 and 28. Four of which are shared with the genetic map. Of these shared hotspots those on linkage group 28 and 15 are shown to a lesser extent using the physical positions than the genetic positions.

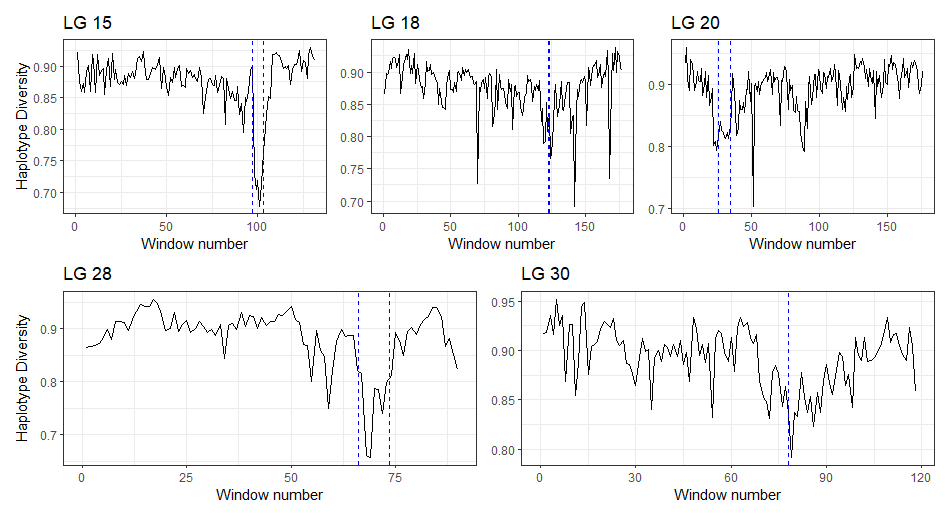


***Haplotype number and diversity in ROH hotspots***

Using ROH hotspots identified using the genetic map shown above we explored haplotype number and diversity in a 10 SNP window. We found no differences in the number of unique haplotypes between ROH hotspots, coldspots or the remainder of the genome. However, when we consider the number of occurrences of each haplotype – i.e. haplotype diversity – we found that that ROH hotspots have a lower diversity than both ROH coldspots and the remainder of the genome (Fig. 3B) (stats). Taken together, the similar number of haplotypes but a lower diversity in ROH hotspots indicates that ROH hotspots have more uneven frequencies in haplotypes than other parts of the genome. From Fig. 4 it is also clear to see a decrease in haplotype diversity in the region that a hotspot occurs compared to the rest of the linkage group. This is particularly illustrated by linkage groups 15 and 28 where the haplotype diversity for the most of the linkage group is relatively stable outside the hotspot region.



**Fig 3** – A: Boxplot of number of unique haplotypes in 10 SNP windows for ROH coldspots, ROH hotspots and the remainder of the genome. B: Boxplot of haplotype diversity for 10 SNP windows in ROH coldspots, ROH hotspots and remainder of the genome. Translucent points show the data points in each group. Black points show the outlier points for each group



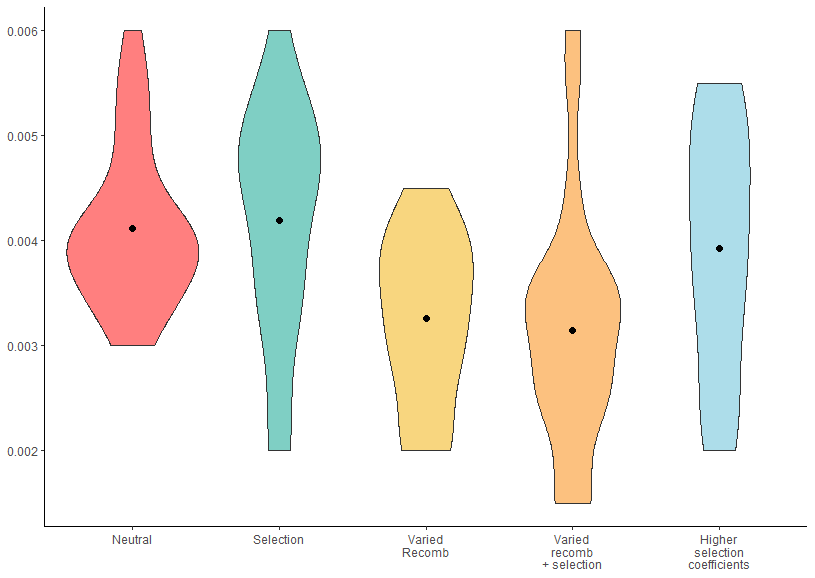
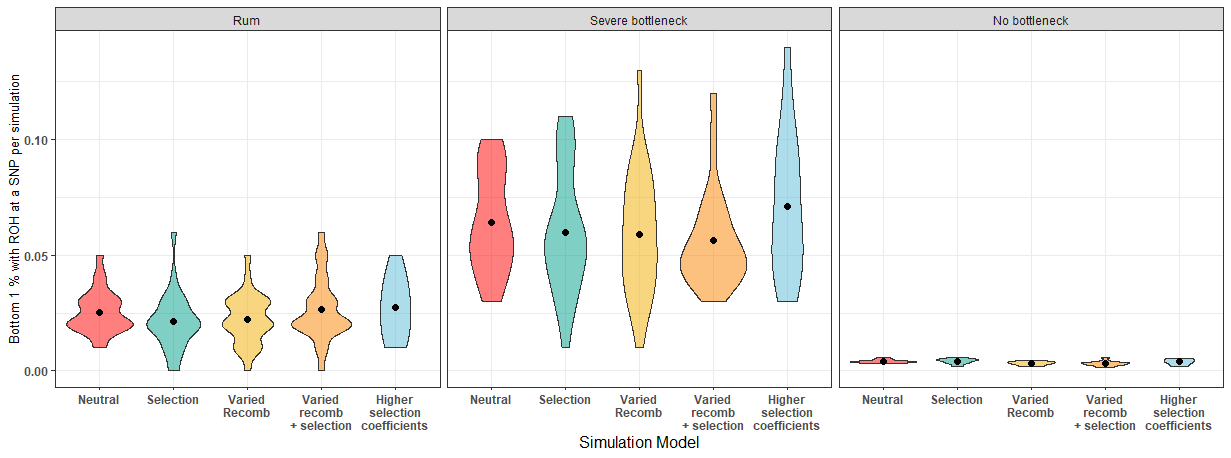
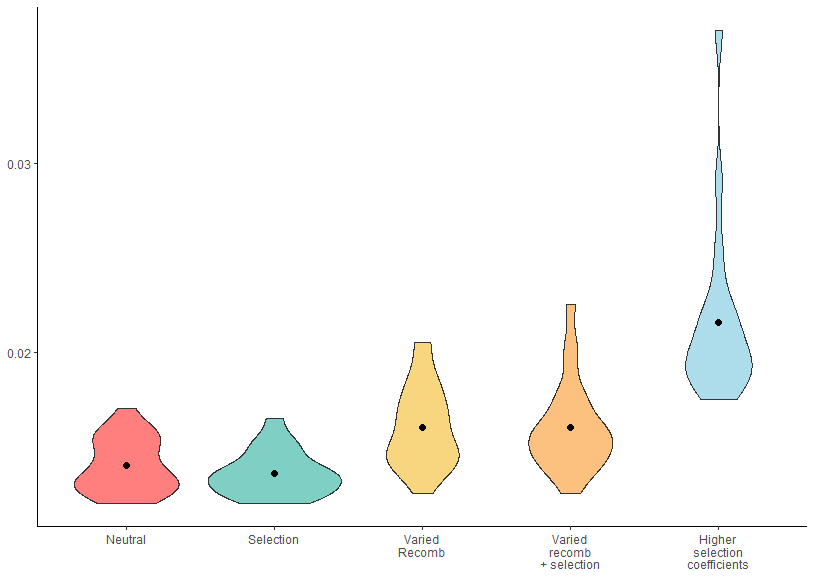
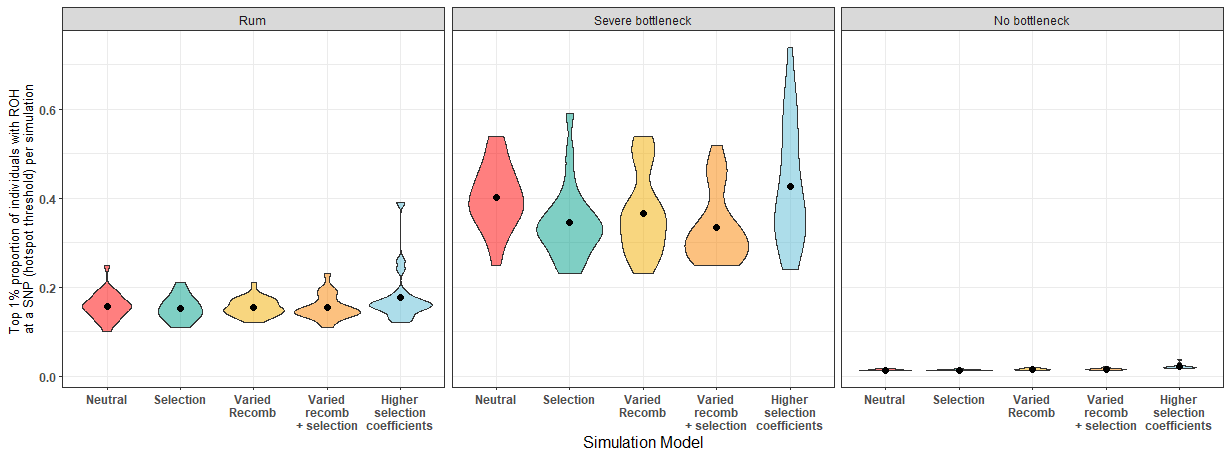
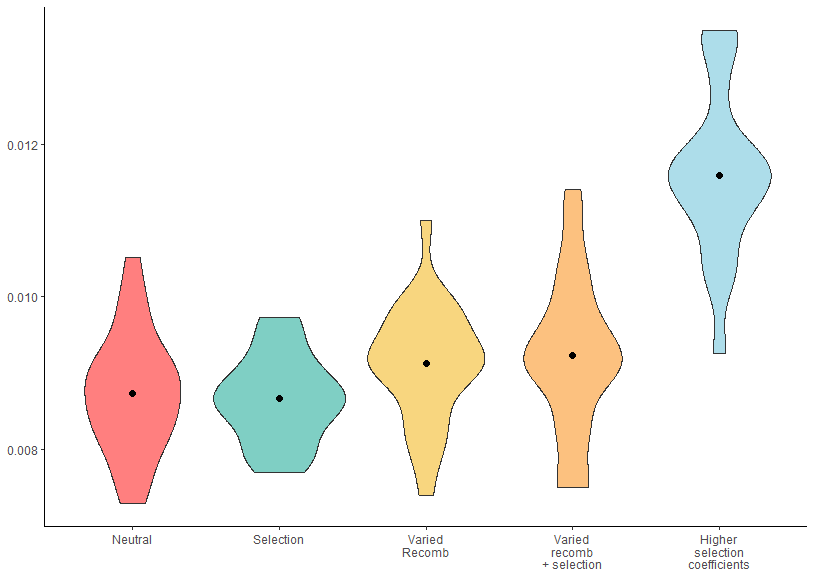
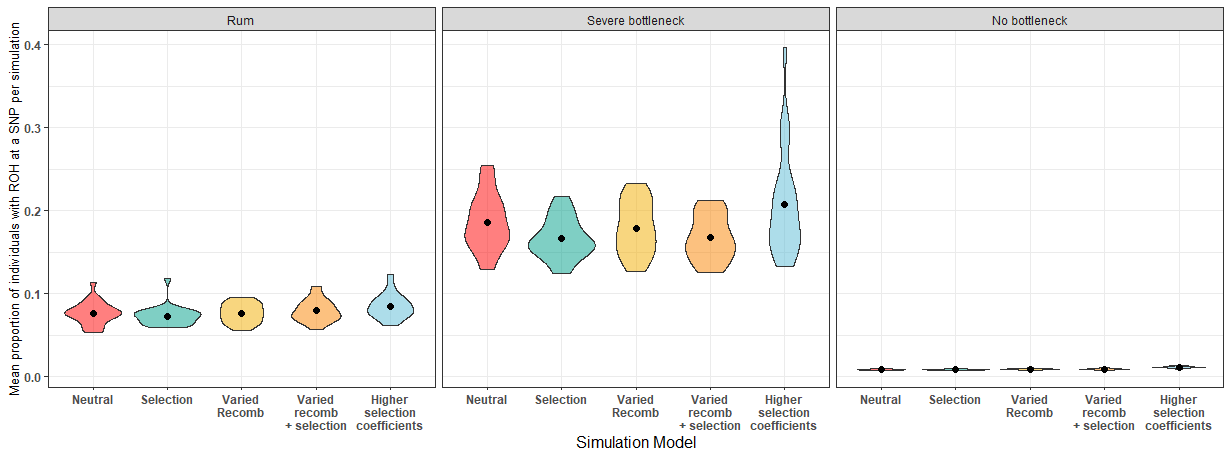
**Fig 4** – Haplotype diversity (D) in 10 SNP windows across linkage groups containing ROH hotspots. Blue dashed lines indicate ROH hotspot regions.

**Population comparison**

***Simulations***

In our simulations, population history had the greatest effect on ROH distributions. For each population history scenario we compared various models to a neutral model with a constant recombination rate and no selection. We investigated on the effects of varying recombination rates over the chromosome, selection, both selection and varied recombination, and higher selection coefficients with varied recombination rates. As can be seen in Fig. within a population history scenario neither model differed greatly from the neutral model for mean *P*. However, when we look between population histories there is a noticeable variation. Having a severe bottleneck gives a higher mean P but with more variability between simulations of the same model. The simulated Rum population history model has a lower mean P. The two models including both varied recombination rates and selection show slightly higher mean P than the neutral model, but this difference is not significant. The population model with no bottleneck understandably has a much lower mean P as it has a higher Ne. The top 1% P (indicating the ROH hotspot threshold) shows a similar pattern between population histories but with slightly more variation within simulation models. For all population histories the variation in mean P and top 1% P varies the most with the higher selection coefficients with varied recombination rate.

When we compare these results to the empirical data from Rum we cannot conclude that the ROH hotspots found are sites of positive selection. In the Rum history simulations mean P and top 1 % P are comparable to the results from the empirical data from Rum. From the simulations we show no model differs greatly from the neutral model. Indicating that the results from the empirical data were equally likely to occur from a neutral model than that of a model which included selection. Equally, we cannot rule out these regions as not sites of selection. However, as the simulations indicate that selection and varied recombination rate does not have a significant effect on the distribution of ROH in a simulated realistic wild population it is unlikely that these ROH hotspot regions found in the empirical data are occurring as a result of selection. (Discussion?)



Discussion

**Population history and inbreeding level**

We found that the Rum population was more inbred that the Kintyre population (Rum mean FROH from genetic map: 0.037, Kintyre FROH from genetic map: 0.019), consistent when using the genetic map or the physical map. Reflecting different population histories from when the Rum population became isolated from the mainland ~150 years ago. The fact that there is more inbreeding in an island population of size 1000 individuals than there is in a much large mainland population is understandable given the lack of dispersal in island populations, which is well documented in literature (REF). Our results for ROH based inbreeding coefficients (FROH) are consistent with other methods of inbreeding coefficients (Fgrm and Fped) reinforcing the observation of inbreeding in these populations. A benefit of using FROH over Fgrm and Fped is the break-down of ROH sizes. By investigating the breakdown of these ROH sizes is it possible to get a clearer outline of a populations history. In the Rum population we found most individuals harboured multiple short and medium ROH. Whereas, very few had long ROH and fewer had multiples of such, which was consistent between methods. This distribution of ROH sizes is comparable to previous studies where short ROH are representative of past inbreeding events from which haplotypes have been broken up into numerous shorter ROH, and long ROH indicate more recent inbreeding events (Peripolli et al., 2018, Pemberton et al., 2012). From outliers in the long ROH in the Rum population it can be inferred that a handful of close inbreeding events have occurred in recent years, supported by findings from Huisman et al. (2016) and what is known through pedigree information. In comparison, the mainland population had a similar distribution of size categories but with much fewer outliers of longer ROH, reflecting the lack of recent inbreeding in a larger population. The pedigree measure of inbreeding coefficients (Fped) was most highly correlated with FROHlong than FROHmedium or FROHshort. Showing that FROHlong, which would reflect recent inbreeding events, can give a good indication of recent population history in populations where for example a pedigree may not be available.

**Recombination rate**

We used two methods when searching for ROH in *plink,* the genetic map reflecting recombination, and the physical map that does not: centimorgan (cM) position, and base pair (BP) position respectively. Kardos et al. (2017) first suggested using the genetic map in this circumstance under the reasoning that if one accounts for recombination, hotspots remaining are more likely to be sites of interest (e.g. areas of positive selection). Kardos et al. (2017) further noted that, when compared to the physical map, the genetic map showed ROH hotspots in the same positions but to a ‘lesser extent’. As in Kardos et al. we find similar lesser hotspots in both populations of red deer. Here, four ROH ‘hotspots’ in the Rum population using the physical map are no longer classed as hotspots when using the genetic map. In addition, we found two hotspots reach further over the hotspot threshold in the genetic map than in the physical map and one new hotspot that reaches over the threshold. Our results support those from Kardos et al. (2017) showing the effect that using the genetic map can have on results analysing ROH hotspots. The differing results are due to the effect that recombination rate has on the distribution of ROH across the genome, which is known from previous studies. Pemberton et al. (2012) etc. clearly show a correlation between ROH distribution and recombination rate, with a number of ROH hotspots occurring in regions of low recombination rate. In using the genetic map positions we can hence account for recombination rate variability across the genome allowing for a better determination of ROH hotspots of interest. Therefore, in agreement with Kardos et al. (2017) we suggest that, where possible, studies should preferentially use genetic map positions to search for ROH hotspots of interest, particularly to infer selection.

ROH distribution is often used to compare populations of the same species and/ or different related species, inferring ROH hotspots common between populations or species are shared sites of selection (refs). Here we find ROH distribution is correlated between two red deer populations in both methods. However, perhaps most noteworthy is that ROH distribution was more highly correlated using the physical map in comparison to the genetic map. As the physical map does not take into account recombination, it suggests that these results are a consequence of recombination rate shared between populations. Nothnagel et al. (2010) also concluded that in humans, locations of ROH islands were alike in different subpopulations due to similar genome-wide patterns of recombination. Similarly, Bosse et al. (2012) shows that pigs from the same population had similar ROH distribution patterns whereas related species had variable ROH distribution resultant from unique population demography events. Our results here support these conclusions, also suggesting recombination rate is driving similarities in ROH patterns between the two populations. In addition, we suggest caution over concluding that ROH hotspots in common between populations are sites of positive selection without taking into account recombination. As demonstrated here, common ROH hotspots may be regions of lower recombination rate, which are common between populations as a result of a shared recombination landscape. Hence, a further benefit of using a genetic map, is in comparing populations will hold fewer biases, giving more accurate representations of ROH distributions occurring from unique population demography or selection.

**Selection**

Once we account for recombination using the genetic map, the remaining hotspots may be areas of interest. A number of studies suggest ROH hotspots as sites of positive selection, in particular, numerous livestock species have been shown to have selected genes enriched in hotspots (Peripolli et al., 2018). When using the genetic map positions the red deer population on Rum showed five genome regions with a ROH in more than 15% of the population i.e. a ROH hotspot. In the mainland population from Kintyre four regions were present in more than 8% of the population. One ROH hotspot was common between the two populations. It is possible these regions may be sites of positive selection, but more analysis is needed to validate this. In ROH hotspots in the Rum population, haplotype diversity was lower in comparison to ROH coldspots and the remainder of the genome. However, the number of different haplotypes did not differ between these groups. Additionally, within specific linkage groups, haplotype diversity was reduced at the site of a hotspot. Together these results suggest certain haplotypes are being favoured at a higher frequency than others. These haplotypes may then be beneficial, representing sites of positive selection. However, inferring sites of selection should be taken in caution. The percentage of individuals with a ROH at a hotspot was much lower here than in other studies showing sites of selection, >80% in livestock and ~20% in humans (Pemberton et al., 2012, Purfield et al., 2012). Therefore, as there is a fairly low number of individuals in the whole population, these may be regions of genetic drift occurring by chance.

**Simulations**

We used simulations as a tool to establish whether the ROH hotspots found in the Rum population may be a result of selection or other means. In the Rum population history simulations, the mean P and top 1% P are similar to the empirical results from Rum, giving confidence that the simulation parameters chosen can reflect the Rum population with some accuracy. From the simulations we show no model in the Rum population history scenario differs greatly from the neutral model. Indicating that the results from the empirical data were equally likely to occur from a neutral model with no selection than that of a model which included selection. However, we cannot conclude these regions are not sites of selection. One noteworthy result common across all population history scenarios is that the model including higher selection coefficients (5x higher than selection coefficients used in other models) resulted in some outliers with a higher top 1% threshold. These simulations with a higher threshold indicate regions of stronger selection driving up this threshold. These outliers were more similar to results from artificially selected organisms, which coincidently have higher selection coefficients than naturally selected organisms. Therefore, ROH hotspots may only contain sites of selection if the selection coefficients for the population were high enough and only in a certain number of outlier cases. In the case of the Rum population, the simulations suggest that selection coefficients may not be high enough to influence ROH distribution. On the other hand, populations with higher selection coefficients (e.g. artificially selected organisms) the selection can drive ROH distribution.

From our simulations, we show that of all the variables tested, population history actually had the greatest influence on ROH distribution. Between population history scenarios there is a noticeable difference in both mean P and top 1% P. The simulations with a severe bottleneck had the highest number of ROH and each model had considerable variability per simulation shown in the spread of the violin plot. These results reflect those observed in real bottlenecked populations. A smaller effective population size means inbreeding will be more common than in populations with no bottleneck, resulting in more ROH in the population. On the other hand, simulations with no bottleneck (i.e. reflecting the mainland population) show a very low incidence of ROH when on the same scale as the other population scenarios. Reflecting the differences found in the empirical data between the mainland and Rum population. The mainland population, with a higher effective population size, will have a lower level of inbreeding resulting in less ROH. We also see a significantly higher mean P and top 1% P in the stronger selection coefficient model relative to the other models in this population history scenario, a pattern which is not present in the other two population history scenarios. This again reflects what is found in literature, than in larger populations selection is more efficient. As such, mutations with higher selection coefficients are likely occurring at a higher frequency and therefore are able to be selected for or against, resulting in a higher incidence of ROH compared to the other models. In a smaller population these mutations may not reach fixation as a result of genetic drift (REF!!).

Our simulations have offered a valuable way to assess how different factors may be affecting our main study population on Rum. Overall showing that population history has the greatest effect on ROH distribution and density in a simulated wild population. Interestingly selection and recombination do not appear to influence the distribution of ROH in our simulations. Selection only appears to have an effect when the selection coefficients are increased to possibly unrealistic amounts for a wild population. In a slightly contrary outcome to our results above and past literature showing the effects of recombination rate on ROH distribution, a model including recombination rate does not differ greatly to a null model. It may be that recombination rate does effect ROH distribution but not to a such dramatic state that would result in a ROH hotspot ??. We have also used rather basic measurements of comparison, as we have 25 iterations of each simulation which may be losing some information.

BJELLAND, D. W., et al. (2013) Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. *J Dairy Sci,* **96** (7), 4697-706.

BOSSE, M., et al. (2012) Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. *PLoS Genet,* **8** (11), e1003100.

CHARLESWORTH, B. (2009) Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat Rev Genet,* **10** (3), 195-205.

GRILZ-SEGER, G., et al. (2019) Analysis of ROH patterns in the Noriker horse breed reveals signatures of selection for coat color and body size. *Anim Genet,* **50** (4), 334-346.

KARDOS, M., et al. (2017) Inferring Individual Inbreeding and Demographic History from Segments of Identity by Descent in Ficedula Flycatcher Genome Sequences. *Genetics,* **205** (3), 1319-1334.

KELLER, M. C., et al. (2012) Runs of homozygosity implicate autozygosity as a schizophrenia risk factor. *PLoS Genet,* **8** (4), e1002656.

KELLER, M. C., et al. (2011) Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics,* **189** (1), 237-49.

NOTHNAGEL, M., et al. (2010) Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. *Hum Mol Genet,* **19** (15), 2927-35.

PEMBERTON, T. J., et al. (2012) Genomic patterns of homozygosity in worldwide human populations. *Am J Hum Genet,* **91** (2), 275-92.

PERIPOLLI, E., et al. (2018) Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gyr (Bos indicus) dairy cattle. *BMC Genomics,* **19** (1), 34.

PRYCE, J. E., et al. (2014) Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle. *Genet Sel Evol,* **46**, 71.

PURFIELD, D. C., et al. (2012) Runs of homozygosity and population history in cattle. *BMC Genet,* **13**, 70.